

Concept Note - Cloud Archival for Master Files

Description

The current size of master files / raw data is around 160 TB, and it is growing around 50 TB annually. These files are in TIFF format. Currently there are 40 hard disks with these master files. Therefore, Cloud shall be the best option to store those files with global access. Noolaham is open to choose any Cloud providers who could serve us well in a cost effective manner.

Master files are kept in 3 different locations at present. Noolaham is willing to explore the backup options in the cloud as well in addition to existing backups. As most of the cloud companies are providing High Availability Disaster Recovery backup solutions, we hope additional backups may not be applicable for Cloud.

Challenges

As the current size of master files are very huge, we need to identify a viable solution for initial upload for the master files. As the current size is 160 TB, it may take longer by uploading using standard internet for initially. As the Noolaham team is actively scanning new documents and adding in to archival storage , those master files must be uploaded into Cloud very frequently or real time. The cloud package should include real time upload access as well.

Noolaham is using TIFF format to scan the original documents for its lossless compression as per Digital Library preservation standards. then these documents will be converted as PDF files for users to view via Noolaham websites. Average size of a TIFF file will be around 30 - 35 MB.

Options

There are few Cloud storage options were analyzed as provided in the table below

	S3 Standard	Glacier	Glacier Deep Archive	Wasabi	Digital Ocean	Google	Comments
Size	200 TB						Current size of Archival documents is 160 TB Around 40 TB data is getting added every year
Document Retrieval	few Seconds	within 5 minutes	12 hours	Few Seconds	Few Seconds		Noolaham is rarely retrieving documents from the Hard Disk and 12 hours waiting time will not be the problem.

Backup	3 Availability Zones			Based on the request	Based on the request		AWS replicates the data storage in 3 Locations
Data Transfer using devices (one time charge)	\$1500 (Not available in Sri Lanka and shall be copied from UK/Australia / Canada)			Wasabi ball is only available with in USA	Not Applicable		A Snowball Device can copy 80 TB data and it would cost around \$500. As this feature is not available in Sri Lanka, Datasync option shall be considered
Data sync (if Data Transfer via Device is not possible)	\$2,500			Wasabi Direct shall be used for minimum 1 year \$1600	SFTP shall be provided		
Monthly Charges	\$4,966	\$1,024	\$412	\$1150 Long term contract may have discount options	\$4,000	~ \$550 (need more research)	
Annual cost	\$59,592	\$12,288	\$4,944	\$13,800	\$48,000		
Non Profit discount	\$1,000						AWS can offer maximum \$1000 credit per year through https://www.techsoup.org/products/amazon-web-services-credits-for-nonprofits-g-50197-

Conclusion

Based on the options discussed above, following 2 options could be cost effective and appropriate for our needs

1. AWS S3 Glacier Deep Archive
2. Wasabi

We need to ensure with them regarding the Retrieval process in case the whole 200TB data needs to be pulled out for any reasons. also, Wasabi could provide more discounts if we go for long term contracts for 1 to 5 years with upfront payment.

